



# UNITED STATES PATENT AND TRADEMARK OFFICE

faw

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
[www.uspto.gov](http://www.uspto.gov)

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/603,212	06/25/2003	Doh-Suk Kim		3320
7590	03/14/2006		EXAMINER	
Docket Administrator (Room 3J-219) Lucent Technologies Inc. 101 Crawfords Corner Road Holmdel, NJ 07733-3030			STORM, DONALD L	
			ART UNIT	PAPER NUMBER
				2654

DATE MAILED: 03/14/2006

Please find below and/or attached an Office communication concerning this application or proceeding.

<b>Office Action Summary</b>	<b>Application No.</b>	<b>Applicant(s)</b>	
	10/603,212	KIM, DOH-SUK	
	<b>Examiner</b> Donald L. Storm	<b>Art Unit</b> 2654	

**— The MAILING DATE of this communication appears on the cover sheet with the correspondence address —**

#### **Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

#### **Status**

- 1) Responsive to communication(s) filed on June 25, 2003 thorough January 28, 2005.
- 2a) This action is FINAL.                    2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

#### **Disposition of Claims**

- 4) Claim(s) 1-20 is/are pending in the application.
  - 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) Claim(s) \_\_\_\_\_ is/are allowed.
- 6) Claim(s) 1-20 is/are rejected.
- 7) Claim(s) \_\_\_\_\_ is/are objected to.
- 8) Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

#### **Application Papers**

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 02 October 2003 is/are: a) accepted or b) objected to by the Examiner.
 

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

#### **Priority under 35 U.S.C. § 119**

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
  - a) All    b) Some \* c) None of:
    1. Certified copies of the priority documents have been received.
    2. Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
    3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

#### **Attachment(s)**

1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892)	4) <input type="checkbox"/> Interview Summary (PTO-413)
2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)	Paper No(s)/Mail Date. _____
3) <input checked="" type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____	5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)
	6) <input checked="" type="checkbox"/> Other: <u>1, 2, 3</u> .

## **DETAILED ACTION**

### ***Allowable Subject Matter***

1. Claims 12, 15-17, and 19-20 would be allowable over the prior art of record if rewritten to include all of the limitations of the base claim and any intervening claims. The whole structure and interaction expressed by the combination of all limitations is not made obvious compared to the prior art of record for the whole invention of those dependent claims, particularly with assessing quality of spec or a distorted version of the speech based on comparing articulation power and non-articulation power and comparing quality of the spec and a distorted version of the speech. Certain assumptions that make the limitations clear have been considered for the claims, as described next or elsewhere in this Office action. The claims should also be rewritten to overcome any objections or rejections under 35 U.S.C. 112(2), especially as appearing in this Office action.

### ***Oath/Declaration***

2. The replacement declaration submitted by the Applicant was received on October 3, 2003, and this declaration is substantively acceptable to the Examiner.

### ***Information Disclosure Statement***

3. A copy of the search report of the European Patent Office (04253532.8-2218-) (received January 28, 2005) is present. The search report and its cited documents have been considered by the Examiner.

4. The information disclosure statement filed January 28, 2005 seems to incorrectly indicate that an English translation was provided. Only one (of two) abstract of the WO 02/43051 provided is in English. The Examiner has changed citations on the disclosure statement to correspond to the copies provided. If the changes are acceptable to the Applicant, no action is required. Further submissions should comply with 37 CFR 1.97 and 37 CFR 1.98 as of the date of their submission.

***Drawings***

5. The proposed substitute drawings (5 sheet(s), received October 2, 2003) are present and are now the Figs. 1-6 of record. These drawing sheets are substantively acceptable to the Examiner.

***Specification***

6. A substitute specification including the claims, abstract, and title is required pursuant to 37 CFR 1.125(a) because the nature of the specification currently of record renders it difficult to consider the application. Copies of the Office record of originally filed specification, claims, and abstract are included with this Office action; the difficulty should be apparent. See also US Patent Application Publication 2004/0267523, which purports to be a publication of this application; note that there are differences.

A substitute specification must not contain new matter. The substitute specification must be submitted with markings showing all the changes relative to the immediate prior version of the specification of record. The text of any added subject matter must be shown by underlining the

added text. The text of any deleted matter must be shown by strikethrough except that double brackets placed before and after the deleted characters may be used to show deletion of five or fewer consecutive characters. The text of any deleted subject matter must be shown by being placed within double brackets if strikethrough cannot be easily perceived. An accompanying clean version (without markings) and a statement that the substitute specification contains no new matter must also be supplied. Numbering the paragraphs of the specification of record is not considered a change that must be shown.

The copies of the originally filed specification, claims, and abstract that are included with this Office action might make it easier for the Applicant to comply with requirements for markings and decisions regarding new matter.

Note elsewhere in this Office action and in US Patent Application Publication 2004/0267523 that the proposed replacement drawings are now of record and this Office action does not require that they be provided again in this application.

### ***Claim Informalities***

7. Claim 1, and by dependency claims 2-20, are objected to under 37 CFR 1.75(a) because the meaning of the phrase “a first and second speech signal” (page 25, lines 5-6) needs clarification. This phrase recites only one signal. However, in order to provide antecedence for each of “the first speech signal” and “the second speech signal”, the Examiner has interpreted this phrase as --first and second speech signals--.

8. Claim 1, and by dependency claims 2-20, are objected to under 37 CFR 1.75(a) because the meaning of the phrase “the first and second speech qualities” (page 25, line 8) needs clarification. Because no speech qualities were previously recited, it may be unclear as to what element this phrase refers. There may be confusion with “first and second speech signals”, because each signal might somehow have quality, or with the singular assessment, because the quality assessment might somehow have quality. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --first and second speech qualities--.

9. Claim 2 is objected to under 37 CFR 1.75(a) because the meaning of the phrase “the additional steps” (page 25, line 11) needs clarification. Because only one additional step is recited by claim 2, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the additional step--.

10. Claim 2 is objected to under 37 CFR 1.75(a) because the meaning of the phrase “the first and second speech quality assessments” (page 25, line 12) needs clarification. Because only one assessment was previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the first and second speech quality assessment--.

However, the Examiner believes that the inventions of both claim 1 and claim 2 might more closely reflect the invention disclosed in the specification and in Fig. 1 if claim 1 were amended to recite “first and second speech quality assessments”, rather than the Examiner’s assumption of a single assessment for claim 2. Note also “the step assessing” in claim 7 and “the step of assessing” in claim 14.

11. Claim 7, and by dependency claims 8-20, are objected to under 37 CFR 1.75(a) because the meaning of the phrase “the step assessing the second or first speech quality” (page 25, lines

27-28) needs clarification. Because no step assessing was previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the step of determining--. Note also “the step of assessing” in claim 14.

12. Claim 7, and by dependency claims 8-20, are objected to under 37 CFR 1.75(a) because the meaning of the phrases “the speech signal or distorted speech signal” (page 25, lines 29-30 and page 26, lines 1-2) needs clarification. Because two speech signals and no distorted speech signal were previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the second speech signal or the distorted version--.

13. Claim 7, and by dependency claims 8-20, are objected to under 37 CFR 1.75(a) because the meaning of the phrase “the comparison” (page 26, line 3) needs clarification. Because both no comparison was previously recited, but both comparing qualities and comparing powers were previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the comparison of the powers--. Note also “the comparison” in claim 14.

14. Claim 12 is objected to under 37 CFR 1.75(a) because the meaning of the phrase “the ratio” (page 26, line 18) needs clarification. Because no ratio was previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --a ratio--. Note that “a ratio” appears in claim 11, but this claim is not dependent to claim 11.

15. Claim 14 is objected to under 37 CFR 1.75(a) because the meaning of the phrase “the comparison” (page 26, line 28) needs clarification. Because all (three) of a comparison, comparing qualities, and comparing powers were previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the comparison of the powers--.

16. Claim 15 is objected to under 37 CFR 1.75(a) because the meaning of the phrase “the local speech quality” (page 26, line 30) needs clarification. Because no local speech quality was previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --a local speech quality--. Note that “a local speech quality” appears in claim 14, but this claim is not dependent to claim 14. The Applicant also may wish to consider whether “further” is a proper adverb for “determined” in view of whatever amendments may be made to this claim.

17. Claim 16 is objected to for the same reasons as claim 15 because the limitations are recited using obviously similar phrases.

18. Claim 18, and by dependency claims 19-20, are objected to under 37 CFR 1.75(a) because the meaning of the phrase “the speech signal” (page 27, line 12) needs clarification. Because two speech signals were previously recited, it may be unclear as to what element this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --the second speech signal or the distorted version--.

19. Claim 20 is objected to under 37 CFR 1.75(a) because the meaning of the phrase “the plurality of modulation spectrums” (page 27, lines 21-22) needs clarification. Because no plurality of modulation spectrums was previously recited, it may be unclear as to what element

this phrase refers. To further timely prosecution and evaluate prior art, the Examiner has interpreted this phase as --a plurality of modulation spectrums--. Note that “a plurality of modulation spectrums” appears in claim 19, but this claim is not dependent to claim 19.

20. The Examiner notes, without objection, the possibility of informalities in the claims. The Applicant may wish to consider changes during normal review and revision of the disclosure.

In claim 7 (page 26, lines 2-3), should the phrase “and and” be --and--?

***Claim Rejections - 35 USC § 102***

21. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

**Hollier**

22. Claims 1-6 are rejected under 35 U.S.C. 102(b) as being anticipated by Hollier et al. [US Patent 5,799,133].

23. Regarding claim 1, Hollier [at column 6, lines 36-51] describes a method of assessing speech by describing the content and functionality of the recited limitations recognizable as a whole to one versed in the art as the following terminology:

determining a first and second assessment(s) for first and second signals [at column 9, lines 14-19, as generate two measures corresponding to the “good” sample and the distorted sample];

comparing first and second to obtain a compensated assessment [at column 9, lines 15-18, as two measures corresponding to the “good” sample and the distorted sample are then compared to produce an error surface];

the first signal being a distorted version of the second signal [at column 8, lines 44-48, as the corresponding second version has distortion of the “good” signal sample];

the signals are speech signals [at column 8, lines 37-48, as speech segments are to be assessed];

assessments and compensated assessment are speech quality assessments [at column 6, lines 36-39, as the model and parameters are for the purpose of measuring signal quality, i.e. the perceived speech quality].

24. Regarding claim 2, Hollier also describes:

prior to the determining [at column 10, lines 42-52, as the source of the distorted signal may be supplied from a store that is pre-generated];

distorting the second signal to produce the first signal [at column 8, lines 44-48, as the corresponding second version has distortion of the “good” signal sample].

25. Regarding claim 3, Hollier also describes:

the qualities are assessed using an identical technique for objective assessment [at column 8, line 63-column 9, line 1, as the “good” and distorted samples have the same model process applied].

26. Regarding claim 4, Hollier also describes:

the compensated assessment corresponds to a difference between the qualities [at column 9, lines 29-32, as values on the error surface are determined as the difference].

27. Regarding claim 5, Hollier also describes:

the compensated assessment corresponds to a ratio between the qualities [at column 9, lines 54-67, as the error entropy of the distortion (the error distribution) corresponds to the logarithm of the magnitude of the error value (numerator) divided by the error energy distribution (reciprocal, denominator)].

28. Regarding claim 6, Hollier also describes:

the qualities are assessed using auditory-articulatory analysis [at column 6, lines 26-38, as the method for estimating parameters relies on audible features for auditory models and a vocal tract model of how sounds are produced to recognize inaudible differences].

#### *Claim Rejections - 35 USC § 103*

29. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

#### *Hollier and Rothenberg*

30. Claims 7-10, 13-14, and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hollier et al. [US Patent 5,799,133] in view of Rothenberg [US Patent 5,454,375].

31. Regarding claim 7, Hollier describes the included claim elements by dependency as indicated elsewhere in this Office action. Hollier [at column 2, lines 1-5] also describes that the spectral content of speech can be used to analyze other characteristics that are not directly observable.

However, Hollier does not explicitly describe comparing articulation power and non-articulation power, wherein articulation and non-articulation powers are associated with articulation and non-articulation frequencies of the signals, and assessing speech quality based on their comparison.

Like Hollier, Rothenberg [at column 1] describes performing auditory-articulatory analysis using speech spectra, and Rothenberg describes the content and functionality of the recited limitations recognizable as a whole to one versed in the art as the following terminology:

power for the speech signal (of distorted version) [at column 6, lines 13-39, as volume velocity of airflow pressure for voice across the impedance of a membrane, measured in some convenient form];

articulation power and non-articulation power associated with articulation and non-articulation frequencies of the speech signal (or distorted version) [at column 1, lines 36-53, as airflow restricted to low, principally subsonic frequencies, below about 20 Hz that measures articulatory patterns of the voice source and airflow making up the voice to at least about 1000 Hz at higher, acoustic frequencies];

comparing them [at column 7, lines 26-29, as sufficiently such that a value of the high frequency energy value is less than a value of the low frequency energy];

and assessing the second speech quality (or first speech quality) based on their comparison [at column 2, lines 21-31, as the quality of the sound produced in the nature of the voice, including a change in the airflow variables being measured].

As indicated, Rothenberg shows that comparing articulation power and non-articulation power, wherein articulation and non-articulation powers are associated with articulation and non-articulation frequencies of the signals, and assessing speech quality based on their comparison was known to artisans at the time of invention. Since Hollier suggests also modeling inaudible characteristics for speech quality assessment and Rothenberg [at column 2, lines 21-25] also points out that changes in the vocal tract acoustics can distort the voice, it would have been

obvious to one of ordinary skill in the art of auditory articulatory analysis at the time of invention to include the concepts described by Rothenberg, at least including comparing articulation power and non-articulation power, wherein articulation and non-articulation powers are associated with articulation and non-articulation frequencies of the signals, and assessing speech quality based on their comparison to derive parameters and use Hollier's vocal tract model or spectral parameters because Rothenberg's articulatory analysis and parameters would provide the advantage of specifying Hollier's models for inaudible characteristics with particular relevance to speech quality.

32. Regarding claim 8, Rothenberg also describes:

the articulation frequencies are approximately 2~12.5 Hz [at column 1, lines 36-40, as airflow restricted to low, principally subsonic frequencies, below about 20 Hz that measures articulatory patterns of the voice source].

33. Regarding claim 9, Rothenberg also describes:

the articulation frequencies correspond approximately to a speed of human articulation [at column 1, lines 36-40, as airflow restricted to low, principally subsonic frequencies, below about 20 Hz may determine the rate of lung deflation].

34. Regarding claim 10, Rothenberg also describes:

the non-articulation frequencies are approximately greater than the articulation frequencies [at column 1, lines 36-54, as airflow restricted to low, principally subsonic frequencies, below about 20 Hz that measures articulatory patterns of the voice source and airflow making up the voice (well above 20 Hz)].

35. Regarding claim 13, Rothenberg also describes:

the comparison is a difference between the articulation power and the non-articulation power [at column 6, lines 25-28, as the impedance that the membrane provides is low impedance for audible frequencies and high impedance for low-frequency components].

36. Regarding claim 14, Rothenberg also describes:

the determined speech quality is local [at column 2, lines 26-31, as the quality of the voice, being produced may have a distorted perception for the speaker].

37. Regarding claim 18, Hollier also describes:

filtering the speech signal to obtain a plurality of critical band signals [at column 9, lines 21-22, as conform the signal to the Bark scale].

**Hollier and Rothenberg and Bell**

38. Claim 11 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hollier et al. [US Patent 5,799,133] in view of Rothenberg [US Patent 5,454,375] and Bell, Jr. et al [US Patent 3,971,034].

39. Regarding claim 11, Hollier and Rothenberg describe and make obvious the included claim elements by dependency as indicated elsewhere in this Office action. Rothenberg [at column 6, lines 13-39] also describes that the volume velocity of airflow pressure for voice, measured in some convenient form, provides a convenient measure of inaudible power of speech.

However, Rothenberg does not discuss the details of making the comparison of articulatory energy to non-articulatory energy. In particular, neither Hollier nor Rothenberg explicitly describes that the comparison is a ratio between articulation power and non-articulation power.

Like Rothenberg, Bell [at column 4] discusses the subsonic frequencies that are present in articulated speech. Although Bell does not use the comparison, Bell describes a comparison of airflow, as follows:

the comparison is a ratio [at column 2, lines 47-48, as inspiration-expiration ratio].

As indicated, Bell shows that a comparison ratio of airflow during speech was known to artisans at the time of invention. The system by Rothenberg requires a comparison of airflow values, but merely any comparison from mature technologies to complete Hollier's inaudible-spectrum model. Rothenberg has not disclosed a preferred approach to the comparison according to a design criterion or solution to any stated problem. Since it appears that the use of any comparison of values that is known to artisans would perform to provide Rothenberg's comparison of articulatory power to non-articulatory power. Rothenberg's application of a comparison of airflow would suggest finding a comparison function used in a similar area, which Bell describes as known to artisans for airflow comparison. It would have been obvious to one of ordinary skill in the art of airflow measurement at the time of invention to include the concepts described by Bell, at least a ratio comparison, because that would provide the comparison of airflow values with which Rothenberg's analysis operates to measure articulatory power and non-articulatory power.

### ***Double Patenting***

#### **Application No. 10/186,862**

40. A rejection based on double patenting of the "same invention" type finds its support in the language of 35 U.S.C. 101 which states that "whoever invents or discovers any new and useful process ... may obtain a patent therefor ..." (Emphasis added). Thus, the term "same invention," in this context, means an invention drawn to identical subject matter. See *Miller v. Eagle Mfg. Co.*, 151 U.S. 186 (1894); *In re Ockert*, 245 F.2d 467, 114 USPQ 330 (CCPA 1957); and *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970).

A statutory type (35 U.S.C. 101) double patenting rejection can be overcome by canceling or amending the conflicting claims so they are no longer coextensive in scope. The filing of a terminal disclaimer cannot overcome a double patenting rejection based upon 35 U.S.C. 101.

41. Claims 1-20 are provisionally rejected under 35 U.S.C. 101 as claiming the same invention as that of claims 10-20 of copending Application No. 10/186,862, which has at least one applicant in common with the instant application. This is a provisional double patenting rejection since the conflicting claims have not in fact been patented. Application No. 10/186,862 has been published as US Patent Application Publication 2004/0002857.

**Application Number 10/186,840 and Hollier**

42. The nonstatutory double patenting rejection is based on a judicially created doctrine grounded in public policy (a policy reflected in the statute) so as to prevent the unjustified or improper timewise extension of the "right to exclude" granted by a patent and to prevent possible harassment by multiple assignees. A nonstatutory obviousness-type double patenting rejection is appropriate where the conflicting claims are not identical, but at least one examined application claim is not patentably distinct from the reference claim(s) because the examined application claim is either anticipated by, or would have been obvious over, the reference claim(s). See, e.g., *In re Berg*, 140 F.3d 1428, 46 USPQ2d 1226 (Fed. Cir. 1998); *In re Goodman*, 11 F.3d 1046, 29 USPQ2d 2010 (Fed. Cir. 1993); *In re Longi*, 759 F.2d 887, 225 USPQ 645 (Fed. Cir. 1985); *In re Van Ornum*, 686 F.2d 937, 214 USPQ 761 (CCPA 1982); *In re Vogel*, 422 F.2d 438, 164 USPQ 619 (CCPA 1970); and *In re Thorington*, 418 F.2d 528, 163 USPQ 644 (CCPA 1969).

A timely filed terminal disclaimer in compliance with 37 CFR 1.321(c) or 1.321(d) may be used to overcome an actual or provisional rejection based on a nonstatutory double patenting ground, AND provided the conflicting application or patent is shown to be commonly owned with

this application or claims an invention made as a result of activities undertaken within the scope of a joint research agreement. See 37 CFR 1.130(b).

Effective January 1, 1994, a registered attorney or agent of record may sign a terminal disclaimer. A terminal disclaimer signed by the assignee must fully comply with 37 CFR 3.73(b).

43. Claims 7-11, 13-15, and 17-20 are provisionally rejected on the ground of nonstatutory, obviousness-type double patenting as being unpatentable over claims 1-16 of copending Application Number 10/186,840, which has at least one applicant in common with the instant application, in view of Hollier et al. [US Patent 5,799,133]. Although the conflicting claims are not identical, they are not patentably distinct from each other because a person of ordinary skill in the art would conclude that the invention defined in the claims in issue is an obvious variation of the invention defined in the claims in Application Number 10/186,840.

This is a provisional obviousness-type double patenting rejection because the conflicting claims have not in fact been patented. Application Number 10/186,840 has been published as US Patent Application Publication 2004/0002852.

44. Regarding claim 7, Application Number 10/186,840 claims the recited limitations recognizable as a whole to one versed in the art as the following terminology in claims 1-16:

comparing articulation power and non-articulation power for the speech signal (or distorted speech signal) [at claim 1, as comparing articulation power and non-articulation power for a speech signal];

wherein articulation and non-articulation powers are powers associated with articulation and non-articulation frequencies of the speech signal (or distorted speech signal) [at claim 1, as wherein articulation and non-articulation powers are powers associated with articulation and non-articulation frequencies of the speech signal];

and assessing the second speech quality (or first speech quality) based on the comparison [at claim 1, as assessing speech quality based on the comparison].

However, Application Number 10/186,840 does not explicitly claim determining a first and second speech quality assessment for a first and second speech signal, the first speech signal being a distorted version of the second speech signal; and comparing the first and second speech qualities to obtain a compensated speech quality assessment, as recited in this application's claim 1.

Like Application Number 10/186,840, Hollier [at column 6, lines 28-41] uses inaudible vocal tract differences to characterize and assess speech quality, and Hollier describes:

determining a first and second assessment(s) for first and second signals [at column 9, lines 14-19, as generate two measures corresponding to the "good" sample and the distorted sample];

comparing first and second to obtain a compensated assessment [at column 9, lines 15-18, as two measures corresponding to the "good" sample and the distorted sample are then compared to produce an error surface];

the first signal being a distorted version of the second signal [at column 8, lines 44-48, as the corresponding second version has distortion of the "good" signal sample];

the signals are speech signals [at column 8, lines 37-48, as speech segments are to be assessed];

assessments and compensated assessment are speech quality assessments [at column 6, lines 36-39, as the model and parameters are for the purpose of measuring signal quality, i.e. the perceived speech quality].

As indicated, Hollier shows that determining a first and second speech quality assessment for a first and second speech signal, the first speech signal being a distorted version of the second speech signal; and comparing the first and second speech qualities to obtain a compensated speech quality assessment was known to artisans at the time of invention. Since Hollier also suggests

modeling inaudible characteristics for speech quality assessment, and Hollier [at column 7, lines 47-55] also points out that the comparison between quality of a speech signal and its distorted version will define a definition function that allows derivation of the quality of signals received for communication channels, it would have been obvious to one of ordinary skill in the art of articulation and non-articulation frequencies and powers of speech at the time of invention to include the concepts described by Hollier, at least including determining a first and second speech quality assessment for a first and second speech signal, the first speech signal being a distorted version of the second speech signal; and comparing the first and second speech qualities to obtain a compensated speech quality assessment with the claimed comparison of articulation and non-articulation powers for a speech signal and making the assessment based on their comparison, because the quality of signals can then be measured by a function derived from the comparison.

Claims 2-16 of Application Number 10/186,840 set forth additional limitations that claim 7 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in Application Number 10/186,840's claims 2-16 differ from those in this application's claim 7 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

45. Regarding claim 8, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 2.

46. Regarding claim 9, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 3.

47. Regarding claim 10, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 4.

48. Regarding claim 11, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 5, and claim 6 by dependency. Claim 6 of Application Number 10/186,840 sets forth additional limitations that claim 11 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in Application Number 10/186,840's claim 6 differ from those in this application's claim 11 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

49. Regarding claim 13, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 7.

50. Regarding claim 14, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 8.

51. Regarding claim 15, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 9, and claims 10-11 by dependency. Claims 10-11 of Application Number 10/186,840 set forth additional limitations that claim 15 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in

Application Number 10/186,840's claims 10-11 differ from those in this application's claim 15 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

52. Regarding claim 17, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 13, and claims 14-16 by dependency. Claims 12 and 14-16 of Application Number 10/186,840 set forth additional limitations that claim 17 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in Application Number 10/186,840's claims 13-16 differ from those in this application's claim 17 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

53. Regarding claim 18, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 14, and claims 15-16 by dependency. Claims 12 and 15-16 of Application Number 10/186,840 set forth additional limitations that claim 18 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in Application Number 10/186,840's claims 15-16 differ from those in this application's claim 18 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

54. Regarding claim 19, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 15, and claim 16 by dependency. Claims 12 and 16 of Application Number 10/186,840 set forth additional limitations that claim 19 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in Application Number 10/186,840's claim 16 differ from those in this application's claim 19 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

55. Regarding claim 20, Application Number 10/186,840 also claims the additional limitations as the additional limitations of dependent claim 16. Claims 12 and 15 of Application Number 10/186,840 set forth additional limitations that claim 20 of this application does not explicitly include.

It would have been obvious to one of ordinary skill in the art of computerized speech recognition at the time that the invention was made that the additional claim limitations in Application Number 10/186,840's claim 16 differ from those in this application's claim 20 only by functions that can be eliminated if the effect of the additional functions is unneeded or undesired. If the functionality provided by the additional limitations were not desired, it would have been obvious to eliminate it, and so achieve the advantage of simplifying the processing.

***Conclusion***

56. The following references here made of record are considered pertinent to applicant's disclosure:

Parra [US Patent 5,313,556] describes comparisons of speech features in the infrasonic range of speech to identify and/or confirm the speaker.

Hollier et al. [US Patent 6,035,270] describes speech features due to inaudible vocal tract differences to characterize speech quality.

Hogden [US Patent 6,052,662] describes speech reconstruction using articulator models at subsonic frequencies to provide higher quality reconstructed speech.

Hardy [US Patent 6,246,978] describes choosing and calculating measures that quantify distortion of speech that could not have been articulated in natural speech.

Ghitza et al. [US Patent 6,609,092] describes a distortion measure between two auditory representations of source and processed speech.

57. Any response to this action should be mailed to:

**Mail Stop Amendment**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**or faxed to:**

(571) 273-8300, (for both formal communications intended for entry and for informal or draft communications, but please label informal fax as "PROPOSED" or "DRAFT")

Patent Correspondence delivered by hand or delivery services, other than the USPS, should be addressed as follows and brought to U.S. Patent and Trademark Office, Customer Service Window, **Mail Stop Amendment**, Randolph Building, 401 Dulany Street, Alexandria, VA 22314

58. Any inquiry concerning this communication or earlier communications from the examiner should be directed to Donald L. Storm, of Art Unit 2654, whose telephone number is (571) 272-7614. The examiner can normally be reached on weekdays between 7:00 AM and 3:30 PM Eastern Time. If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on (571) 272-7602.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Inquiries regarding the status of submissions relating to an application or questions on the Private PAIR system should be directed to the Electronic Business Center (EBC) at 866-217-9197 (toll-free) or 571-272-4100 between the hours of 6 a.m. and midnight Monday through Friday EST, or by e-mail at: [ebc@uspto.gov](mailto:ebc@uspto.gov). For general information about the PAIR system, see <http://pair-direct.uspto.gov>.

March 3, 2006

*Donald L. Storm*  
Donald L. Storm  
Examiner, Art Unit 2654

Attachment 3 to Office action, paper number 20060303

— ORIGINALLY FILED ABSTRACT —

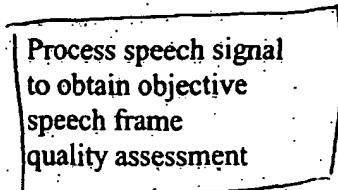
D.S. Kim 3

Abstract of the Disclosure

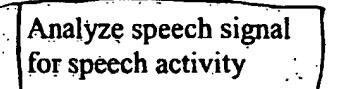
A method for objective speech quality assessment that accounts for phonetic contents, speaking styles or individual speaker differences by distorting speech signals under speech quality assessment. By using a distorted version of a speech signal, 5 it is possible to compensate for different phonetic contents, different individual speakers and different speaking styles when assessing speech quality. The amount of degradation in the objective speech quality assessment by distorting the speech signal is maintained similarly for different speech signals, especially when the amount of distortion of the distorted version of speech signal is severe. Objective speech quality assessment for the 10 distorted speech signal and the original undistorted speech signal are compared to obtain a speech quality assessment compensated for utterance dependent articulation.

100

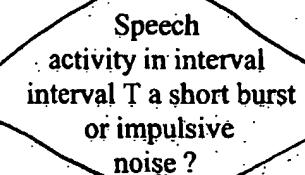
step 102



step 105

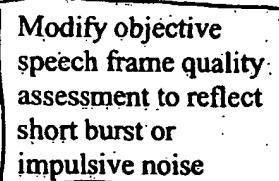


step 110



Yes

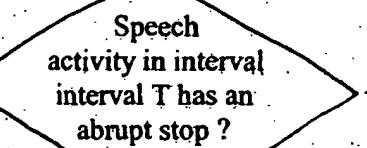
step 115



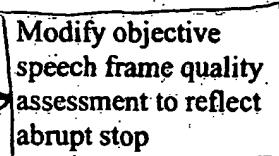
No

step 125

step 120



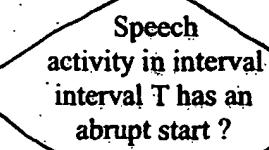
Yes



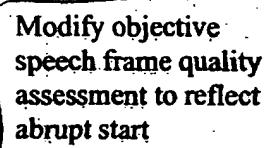
No

step 135

step 130



Yes



No

step 140

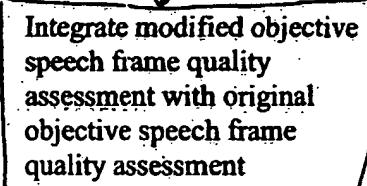


Figure 1

200

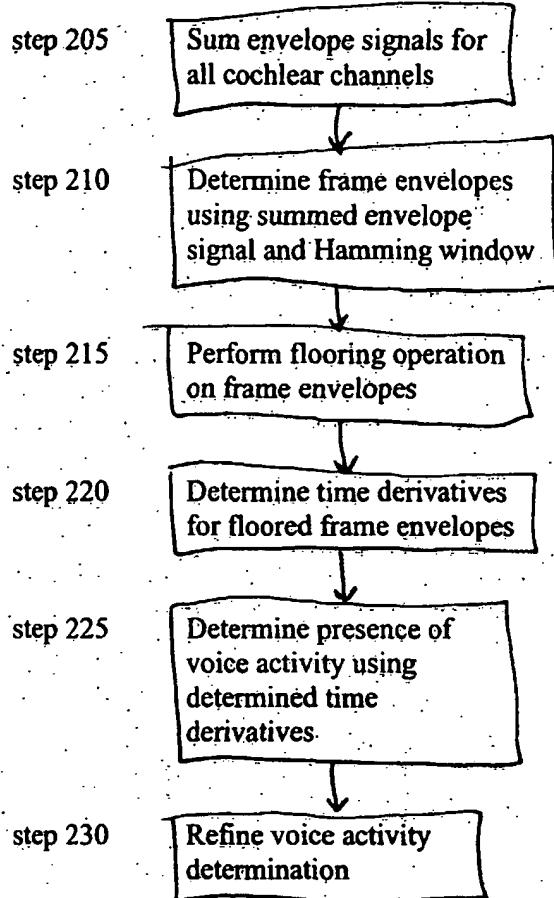


Figure 2

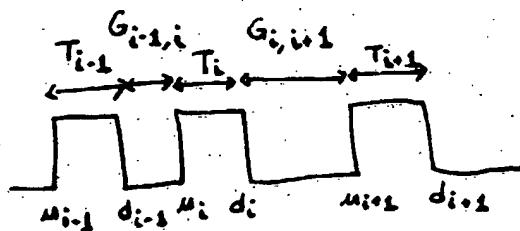


Figure 3

400

step 405

Determine impulsive noise frame  $l_1$  for speech activity interval T

step 410

Frame energy for impulsive noise frame  $l_1 >$  noise threshold ?

Yes

step 415

Speech activity interval T not short burst or impulsive noise

step 420

Interval T  $\geq$  perception threshold and  $\leq$  short burst threshold ?

Yes

step 425

Speech activity interval T not short burst or impulsive noise

step 430

Maximum delta frame envelope  $>$  abrupt change threshold ?

Yes

step 435

Speech activity interval T not short burst or impulsive noise

step 440

Frame  $m_1$  sufficiently annoying to human listener ?

Yes

step 445

Speech activity interval T not short burst or impulsive noise

step 445

Human speech ?

Yes

step 455

Speech activity interval T not short burst or impulsive noise

step 460

Modify objective speech frame quality assessment

Figure 4

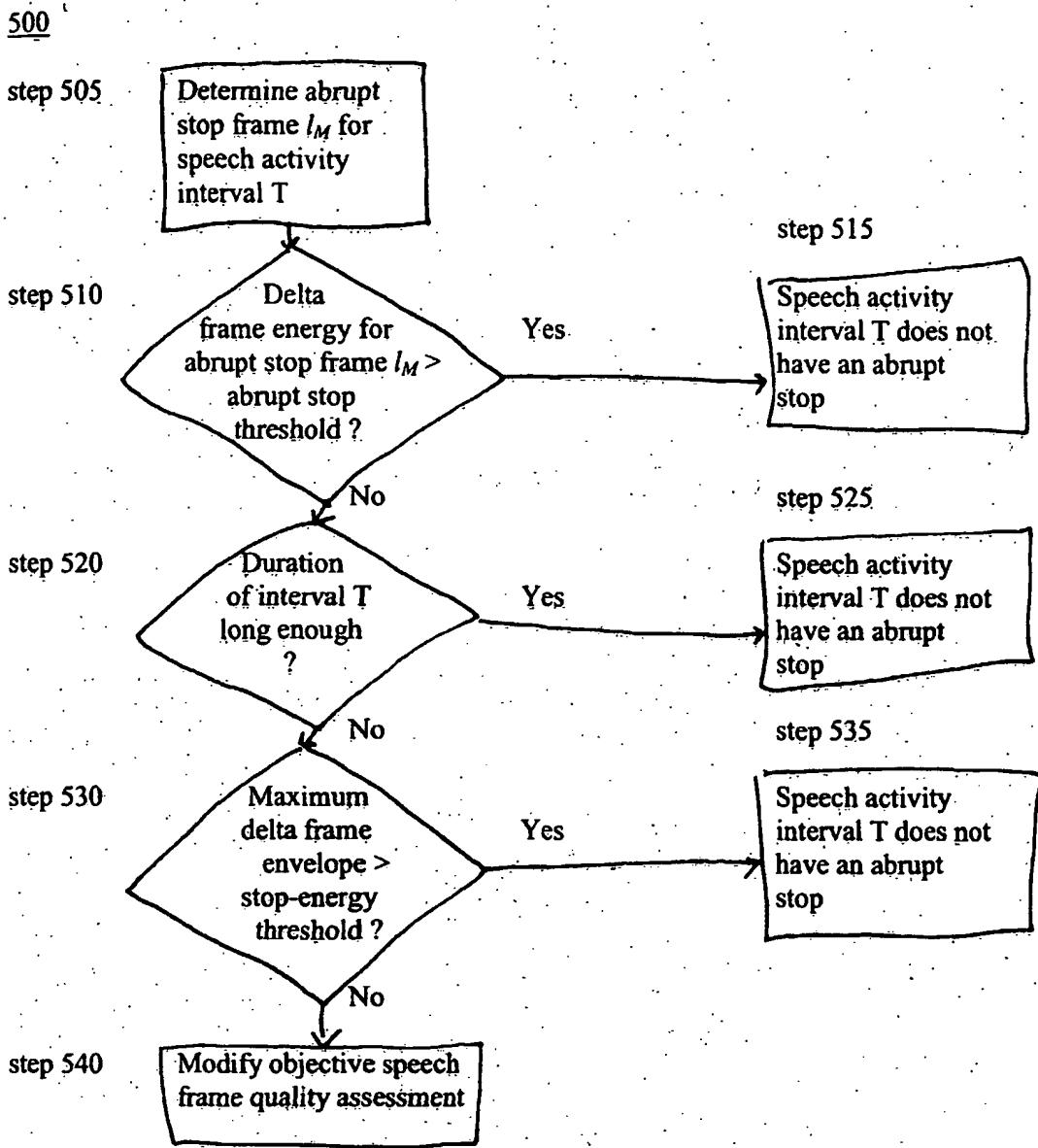


Figure 5

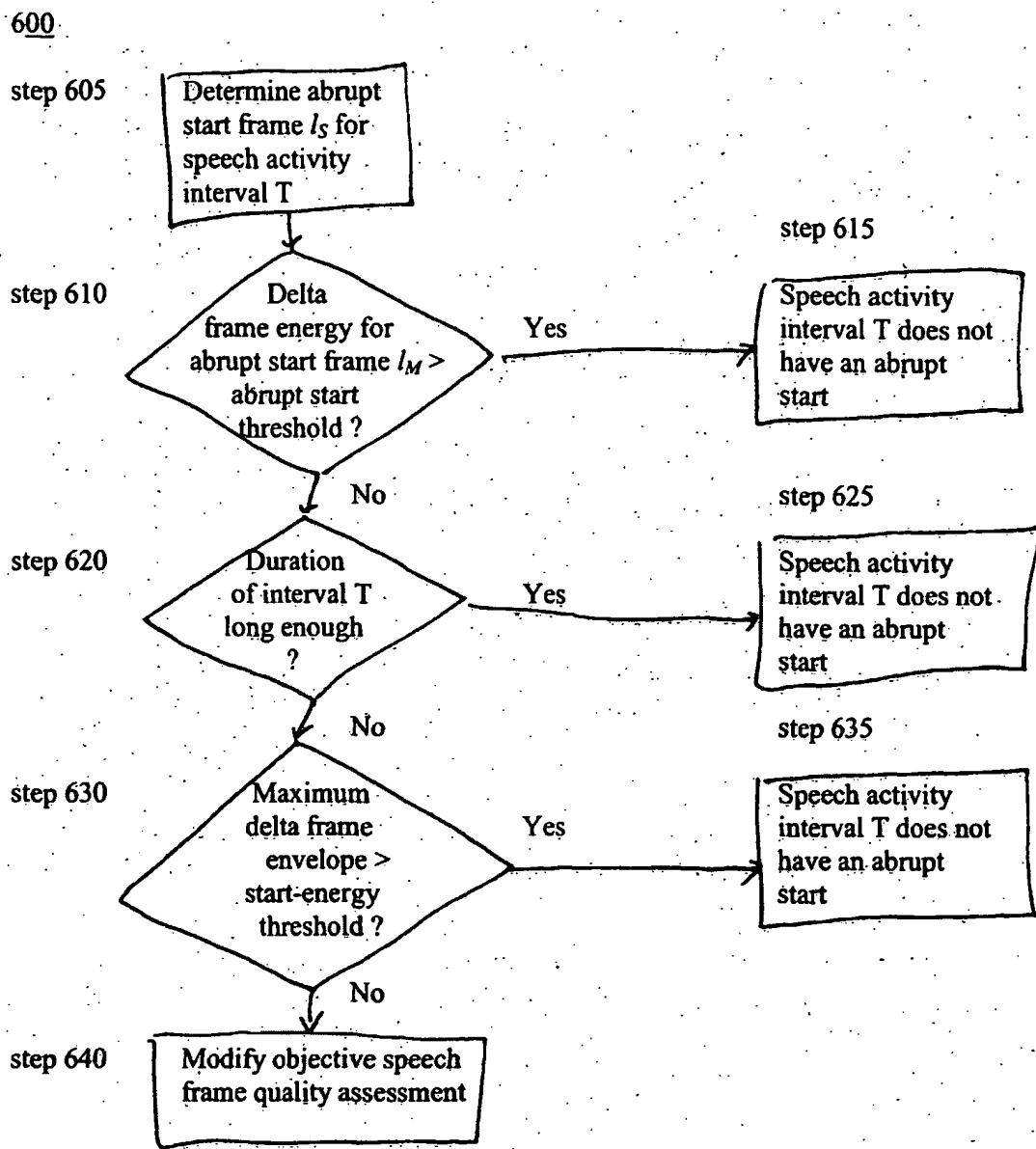


Figure 6

Attachment 2 to Office action, paper number 200060303

— ORIGINALLY FILED CLAIMS —

D.S. Kim 3

Claims

I claim:

1. A method of assessing speech quality comprising the steps of:
  5. determining a first and second speech quality assessment for a first and second speech signal, the first speech signal being a distorted version of the second speech signal; and
    10. comparing the first and second speech qualities to obtain a compensated speech quality assessment.
15. 2. The method of claim 1 comprising the additional steps of prior to determining the first and second speech quality assessments, distorting the second speech signal to produce the first speech signal.
20. 3. The method of claim 1, wherein the first and second speech qualities are assessed using an identical technique for objective speech quality assessment.
25. 4. The method of claim 1, wherein the compensated speech quality assessment corresponds to a difference between the first and second speech qualities.
30. 5. The method of claim 1, wherein the compensated speech quality assessment corresponds to a ratio between the first and second speech qualities.
6. The method of claim 1, wherein the first and second speech qualities are assessed using auditory-articulatory analysis.
7. The method of claim 1, wherein the step assessing the second or first speech quality comprises the steps of;
  - comparing articulation power and non-articulation power for the speech signal or distorted speech signal, wherein articulation and non-articulation powers

are powers associated with articulation and non-articulation frequencies of the speech signal or distorted speech signal; and

and assessing the second or first speech quality based on the comparison.

5        8. The method of claim 7, wherein the articulation frequencies are approximately  
2~12.5 Hz.

9. The method of claim 7, wherein the articulation frequencies correspond  
approximately to a speed of human articulation.

10      10. The method of claim 7, wherein the non-articulation frequencies are  
approximately greater than the articulation frequencies.

11      15. The method of claim 7, wherein the comparison between the articulation power  
and non-articulation power is a ratio between the articulation power and non-  
articulation power.

12      20. The method of claim 10, wherein the ratio includes a denominator and numerator,  
the numerator including the articulation power and a small constant, the  
denominator including the non-articulation power plus the small constant.

13      25. The method of claim 7, wherein the comparison between the articulation power  
and non-articulation power is a difference between the articulation power and non-  
articulation power.

14      30. The method of claim 7, wherein the step of assessing the first or second speech  
quality includes the step of:  
determining a local speech quality using the comparison.

15. The method of claim 7, wherein the local speech quality is further determined  
using a weighing factor based on a DC-component power.

16. The method of claim 9, wherein the first or second speech quality is determined using the local speech quality.
- 5    17. The method of claim 7, wherein the step of comparing articulation power and non-articulation power includes the step of:  
                performing a Fourier transform on each of a plurality of envelopes obtained from a plurality of critical band signals.
- 10    18. The method of claim 7, wherein the step of comparing articulation power and non-articulation power includes the step of:  
                filtering the speech signal to obtain a plurality of critical band signals.
- 15    19. The method of claim 18, wherein the step of comparing articulation power and non-articulation power includes the step of:  
                performing an envelope analysis on the plurality of critical band signals to obtain a plurality of modulation spectrums.
- 20    20. The method of claim 18, wherein the step of comparing articulation power and non-articulation power includes the step of:  
                performing a Fourier transform on each of the plurality of modulation spectrums.

Attachment 1 to Office action, paper number 20060303

— ORIGINALLY FILED SPECIFICATION —

D.S. Kim 4

**METHOD OF REFLECTING TIME/LANGUAGE DISTORTION IN  
OBJECTIVE SPEECH QUALITY ASSESSMENT**

**Field of the Invention**

5. The present invention relates generally to communications systems and, in particular, to speech quality assessment.

**Background of the Related Art**

Performance of a wireless communication system can be measured, 10 among other things, in terms of speech quality. In the current art, there are two techniques of speech quality assessment. The first technique is a subjective technique (hereinafter referred to as "subjective speech quality assessment"). In subjective speech quality assessment, human listeners are typically used to rate the speech quality of processed speech, wherein processed speech is a transmitted speech signal which has 15 been processed at the receiver. This technique is subjective because it is based on the perception of the individual human, and human assessment of speech quality by native listeners, i.e., people that speak the language of the speech material being presented or listened, typically takes into account language effects. Studies have shown that a listener's knowledge of language affects the scores in subjective listening tests. Scores 20 given by native listeners were lower in subjective listening tests compared to scores given by non-native listeners when language information in speech is defect, i.e., mute. In a normal telephone conversation, the listener is often a native listener. Thus, it is preferable to use native listeners for subjective speech quality assessment in order to emulate typical conditions. Subjective speech quality assessment techniques provide a 25 good assessment of speech quality but can be expensive and time consuming.

The second technique is an objective technique (hereinafter referred to as "objective speech quality assessment"). Objective speech quality assessment is not based on the perception of the individual human. Some objective speech quality assessment techniques are based on known source speech or reconstructed source speech estimated 30 from processed speech. Other objective speech quality assessment techniques are not based on known source speech but on processed speech only. These latter techniques are

referred to herein as "single-ended objective speech quality assessment techniques" and are often used when known source speech or reconstructed source speech are unavailable.

Current single-ended objective speech quality assessment techniques, however, do not provide as good an assessment of speech quality compared to subjective speech quality assessment techniques. One reason why current single-ended objective speech quality assessment techniques are not as good as subjective speech quality assessment techniques is because the former techniques do not account for language effects. Current single-ended objective speech quality assessment techniques have been unable to account for language effects in its speech assessment.

10 Accordingly, there exists a need for a single-ended objective speech quality assessment technique which accounts for language effects in assessing speech quality.

Summary of the Invention

15 The present invention is an objective speech quality assessment technique that reflects the impact of distortions which can dominate overall speech quality assessment by modeling the impact of such distortions on subjective speech quality assessment, thereby, accounting for language effects in objective speech quality assessment. In one embodiment, the objective speech quality assessment technique of the  
20 present invention comprises the steps of detecting distortions in an interval of speech activity using envelope information, and modifying an objective speech quality assessment value associated with the speech activity to reflect the impact of the distortions on subjective speech quality assessment. In one embodiment, the objective speech quality assessment technique also distinguish types of distortions, such as short bursts, abrupt stops and abrupt starts, and modifies the objective speech quality  
25 assessment values to reflect the different impacts of each type of distortion on subjective speech quality assessment.

Brief Description of the Drawings

The features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings where:

5 Fig. 1 depicts a flowchart illustrating an objective speech quality assessment technique accounting for language effects in accordance with one embodiment of the present invention;

Fig. 2 depicts a flowchart illustrating a voice activity detector (VAD) which detects voice activity by examining envelope information associated with the speech signal in accordance with one embodiment of the present invention;

10 Fig. 3 depicts an example VAD activity diagram illustrating intervals T and G of speech and non-speech activities, respectively;

Fig. 4 depicts a flowchart illustrating an embodiment for determining whether speech activity is a short burst or impulsive noise and for modifying objective speech frame quality assessment  $v_s(m)$  when a short burst or impulsive noise is determined;

15 Fig. 5 depicts a flowchart illustrating an embodiment for determining whether speech activity has an abrupt stop or mute and for modifying objective speech frame quality assessment  $v_s(m)$  when it is determined that such speech activity has an abrupt stop or mute; and

20 Fig. 6 depicts a flowchart illustrating an embodiment for determining whether speech activity has an abrupt start and for modifying objective speech frame quality assessment  $v_s(m)$  when it is determined that such speech activity has an abrupt start.

Detailed Description

25 The present invention is an objective speech quality assessment technique that reflects the impact of distortions which can dominate overall speech quality assessment by modeling the impact of such distortions on subjective speech quality assessment, thereby, accounting for language effects in objective speech quality assessment.

30 Fig. 1 depicts a flowchart 100 illustrating an objective speech quality assessment technique accounting language effects in accordance with one embodiment of

the present invention. In step 102, speech signal  $s(n)$  is processed to determine objective speech frame quality assessment  $v_s(m)$ , i.e., objective quality of speech at frame  $m$ . In one embodiment, each frame  $m$  corresponds to a 64 ms interval. The manner of processing a speech signal  $s(n)$  to obtain objective speech frame quality assessment  $v_s(m)$  (which do not account for language effects) is well-known in the art. One example of such processing is described in co-pending application serial number 10/186,862, entitled "Compensation Of Utterance-Dependent Articulation For Speech Quality Assessment", filed on July 01, 2002 by inventor Doh-Suk Kim, attached herein as Appendix A.

In step 105, speech signal  $s(n)$  is analyzed for voice activity by, for example, a voice activity detector (VAD). VADs are well-known in the art. Fig. 2 depicts a flowchart 200 illustrating a VAD which detects voice activity by examining envelope information associated with the speech signal in accordance with one embodiment of the present invention. In step 205, envelope signals  $\gamma_k(n)$  are summed up for all cochlear channels  $k$  to form summed envelope signal  $\gamma(n)$  in accordance with equation (1):

$$\gamma(n) = \sum_{k=1}^{N_{cb}} \gamma_k(n) \quad \text{equation (1)}$$

where  $\gamma_k(n) = \sqrt{s_k^2(n) + \hat{s}_k^2(n)}$ ,  $n$  represents a time index,  $N_{cb}$  represents a total number of critical bands,  $s_k(n)$  represents the output of speech signal  $s(n)$  through cochlear channel  $k$ , i.e.,  $s_k(n) = s(n) * h_k(n)$ , and  $\hat{s}_k(n)$  is the Hilbert transform of  $s_k(n)$ .

In step 210, a frame envelope  $e(l)$  is computed every 2 ms by multiplying summed envelope signal  $\gamma(n)$  with a 4 ms Hamming window  $w(n)$  in accordance with equation (2):

$$e(l) = \log \left[ \sum_{n=0}^{31} \gamma^{(l)}(n) w(n) + 1 \right] \quad \text{equation (2)}$$

where  $\gamma^{(l)}(n)$  is the 2 ms  $l$ -th frame signal of the summed envelope signal  $\gamma(n)$ . It should be understood that the durations of the frame envelope  $e(l)$  and Hamming window  $w(n)$  are merely illustrative and that other durations are possible. In step 215, a flooring operation is applied to frame envelope  $e(l)$  in accordance with equation (3).

$$e(l) = \begin{cases} e(l) & \text{if } e(l) > 5 \\ 5 & \text{otherwise} \end{cases} \quad \text{equation (3)}$$

In step 220, time derivative  $\Delta e(l)$  of floored frame envelope  $e(l)$  is obtained in accordance with equation (4).

$$\Delta e(l) = \frac{\sum_{j=-3}^3 j e(l-j)}{\sum_{j=-3}^3 j^2} \quad \text{equation (4)}$$

5 where  $-3 \leq j \leq 3$ .

In step 225, voice activity detection is performed in accordance with equation (5).

$$vad(l) = \begin{cases} 1 & \text{if } e(l) > 5 \\ 0 & \text{otherwise} \end{cases} \quad \text{equation (5)}$$

In step 230, the result of equation (5), i.e.,  $vad(l)$ , can then be refined based on the duration of 1's and 0's in the output. For example, if the duration of 0's in  $vad(l)$  is shorter than 8 ms, then  $vad(l)$  shall be changed to 1's for that duration. Similarly, if the duration of 1's in  $vad(l)$  is shorter than 8 ms, the  $vad(l)$  shall be changed to 0's for that duration. Fig. 3 depicts an example VAD activity diagram 30 illustrating intervals T and G of speech and non-speech activities, respectively. It should be understood that speech activities associated with intervals T may include, for example, actual speech, data or noise.

Returning to flowchart 100 of Fig. 1, upon analyzing speech signal  $s(n)$  for speech activity, interval T is examined to determine whether the associated speech activity corresponds to a short burst or impulsive noise in step 110. If the speech activity in interval T is determined to be a short burst or impulsive noise, then objective speech frame quality assessment  $v_s(m)$  is modified in step 115 to obtain a modified objective speech frame quality assessment  $\hat{v}_s(m)$ . The modified objective speech frame quality assessment  $\hat{v}_s(m)$  accounts for the effects of short burst or impulsive noise by modeling or simulating the impact of short bursts or impulsive noise on subjective speech quality assessment.

From step 115 or if in step 110 the speech activity in interval T is not determined to be a short burst or impulsive noise, then flowchart 100 proceeds to step 120 where the speech activity in interval T is examined to determine whether it has an abrupt stop or mute. If the speech activity in interval T is determined to have an abrupt stop or mute, then objective speech frame quality assessment  $v_s(m)$  is modified in step 125 to obtain a modified objective speech frame quality assessment  $\hat{v}_s(m)$ . The modified objective speech frame quality assessment  $\hat{v}_s(m)$  accounts for the effects of the abrupt stop or mute by modeling or simulating the impact of an abrupt stop or mute and subsequent release on subjective speech quality assessment.

From step 125 or if in step 120 the speech activity in interval T is not determined to have an abrupt stop or mute, then flowchart 100 proceeds to step 130 where the speech activity in interval T is examined to determine whether it has an abrupt start. If the speech activity in interval T is determined to have an abrupt start, then objective speech frame quality assessment  $v_s(m)$  is modified in step 135 to obtain a modified objective speech frame quality assessment  $\hat{v}_s(m)$ . The objective speech frame quality assessment  $v_s(m)$  accounts for the effects of the abrupt start by modeling or simulating the impact of an abrupt start on subjective speech quality assessment. From step 135 or if in step 130 the speech activity in interval T is not determined to have an abrupt start, then flowchart 100 proceeds to step 145 where the results of modifications to objective speech frame quality assessment  $v_s(m)$ , if any, are integrated into the original objective speech frame quality assessment  $v_s(m)$  of step 102.

Techniques for determining whether speech activity is a short burst (or impulsive noise) or has an abrupt stop (or mute) or an abrupt start, i.e., steps 110, 120 and 130, along with techniques for modifying objective speech frame quality assessment  $v_s(m)$ , i.e., steps 115, 125 and 135, in accordance with one embodiment of the invention will now be described. Fig. 4 depicts a flowchart 400 illustrating an embodiment for determining whether speech activity is a short burst or impulsive noise and for modifying objective speech frame quality assessment  $v_s(m)$  when a short burst or impulsive noise is determined. In step 405, an impulsive noise frame  $I_I$  is determined by finding a frame  $I$  in interval  $T_I$  where frame envelope  $e(I)$  is maximum in accordance, for example, with equation (6):

$$l_i = \arg \max_{u_i \leq l \leq d_i} e(l) \quad \text{equation (6)}$$

where  $u_i$  and  $d_i$  represents frames  $l$  at the beginning and end of interval  $T_i$ , respectively.

In step 410, frame envelope  $e(l_i)$  is compared to a listener threshold value indicating whether a human listener can consider the corresponding frame  $l_i$  as annoying short burst.

5 In one embodiment, the listener threshold value is 8 — that is, in step 410,  $e(l_i)$  is checked to determine whether it is greater than 8. If frame envelope  $e(l_i)$  is not greater than the listener threshold value, then in step 415 the speech activity is determined not to be a short burst or impulsive noise.

If frame envelope  $e(l_i)$  is greater than the listener threshold value, then in  
 10 step 420 the duration of interval  $T_i$  is checked to determine whether it satisfies both a short burst threshold value and a perception threshold value. That is, interval  $T_i$  is being checked to determine whether interval  $T_i$  is not too short to be perceived by a human listener and not too long to be categorized as a short burst. In one embodiment, if the duration of interval  $T_i$  is greater than or equal to 28 ms and less than or equal to 60 ms,  
 15 i.e.,  $28 \leq T_i \leq 60$ , then both of the threshold values of step 420 are satisfied. Otherwise the threshold values of step 420 are not satisfied. If the threshold values of step 420 are not satisfied, then in step 425 the speech activity is determined not to be a short burst or impulsive noise.

If the threshold values of step 420 are satisfied, then in step 430 a  
 20 maximum delta frame envelope  $\Delta e(l)$  is determined from the frame envelopes  $e(l)$  in the one or more frames prior to the beginning of interval  $T_i$  through the first one or more frames of interval  $T_i$  and subsequently compared to an abrupt change threshold value, such as 0.25. The abrupt change threshold value representing a criteria for identifying an abrupt change in the frame envelope. In one embodiment, a maximum delta frame envelope  $\Delta e(l)$  is determined from frame envelope  $e(u_i-1)$ , i.e., frame envelope immediately preceding interval  $T_i$ , through the frame envelope  $e(u_i+5)$ , i.e., fifth frame envelope in interval  $T_i$ , and compared to a threshold value of 0.25 — that is, in step 430, it is checked to determine whether equation (7) is satisfied:

$$\max_{u_i-1 \leq l \leq u_i+5} \Delta e(l) > 0.25 \quad \text{equation (7)}$$

If the maximum delta frame envelope  $\Delta e(l)$  does not exceed the threshold value, then in step 435 the speech activity is determined not to be a short burst or impulsive noise.

If the maximum delta frame envelope  $\Delta e(l)$  does exceed the threshold value, then in step 440 it is determined whether frame  $m_l$  would be sufficiently annoying

5 to a human listener, where  $m_l$  corresponds to the frame  $m$  which is impacted most by impulsive noise frame  $l$ . In one embodiment, step 440 is achieved by determining whether a ratio of objective speech frame quality assessment  $v_s(m_l)$  to modulation noise reference unit  $v_q(m_l)$  exceeds a noise threshold value. Step 440 may be expressed, for example, using a noise threshold value of 1.1 and equation (8):

$$10 \quad \frac{v_s(m_l)}{v_q(m_l)} < 1.1 \quad \text{equation (8)}$$

wherein if equation (8) is satisfied, it would be determined that frame  $m_l$  has sufficient annoyance to a human listener. If it is determined that objective speech frame quality assessment  $v_s(m_l)$  would be sufficiently annoying to a human listener, then in step 445 the speech activity is determined not to be a short burst or impulsive noise.

15 If it is determined that objective speech frame quality assessment  $v_s(m_l)$  would not be sufficiently annoying to a human listener, then in step 450 conditions related to the durations of intervals  $G_{i-1,j}$ ,  $G_{i,i+1}$ ,  $T_{i-1}$  and/or  $T_{i+1}$  satisfying certain minimum or maximum duration threshold values are checked to verify that it belongs to human speech. In one embodiment, the conditions of step 450 are expressed as equations 20 (9) and (10).

$$G_{i-1,j} < 180 \text{ ms and } G_{i,i+1} > 40 \text{ ms and } T_{i-1} > 50 \text{ ms} \quad \text{equation (9)}$$

$$G_{i-1,j} > 40 \text{ ms and } G_{i,i+1} < 100 \text{ ms and } T_{i+1} > 60 \text{ ms} \quad \text{equation (10)}$$

If any of these equations or conditions are satisfied, then in step 455 the speech activity is determined not to be a short burst or impulsive noise. Rather the speech activity is 25 determined to be natural speech. It should be understood that the minimum and maximum duration threshold values used in equations (9) and (10) are merely illustrative and may be different.

If none of the conditions in step 450 are satisfied, then in step 460 objective speech frame quality assessment  $v_s(m)$  is modified in accordance with equation

30 11:

$$\theta_3(m) = \frac{v_s(m)}{1 + \exp[-8.2(m - m_i)/e(l_i) - 10]} \quad \text{equation (11)}$$

Fig. 5 depicts a flowchart 500 illustrating an embodiment for determining whether speech activity has an abrupt stop or mute and for modifying objective speech frame quality assessment  $v_s(m)$  when it is determined that such speech activity has an abrupt stop or mute. In step 505, abrupt stop frame  $l_M$  is determined. The abrupt stop frame  $l_M$  is determined by first finding negative peaks of delta frame envelope  $\Delta e(l)$  in the speech activity using all frames  $l$  in interval  $T_i$ . Delta frame envelope  $\Delta e(l)$  has a negative peak at  $l$  if  $\Delta e(l) < \Delta e(l+j)$  for  $3 \leq j \leq 3$ . Upon finding the negative peaks, abrupt stop frame  $l_M$  is determined as the minimum of the negative peaks of delta frame envelopes  $\Delta e(l)$ . In step 510, delta frame envelope  $\Delta e(l_M)$  is checked to determine whether an abrupt stop threshold value is satisfied. The abrupt stop threshold representing a criteria for determining whether there was sufficient negative change in frame envelope from one frame  $l$  to another frame  $l+1$  to be considered an abrupt stop. In one embodiment, the abrupt stop threshold value is -0.56 and step 510 may be expressed as equation (12):

$$\Delta e(l_M) < -0.56 \quad \text{equation (12)}$$

If delta frame envelope  $\Delta e(l_M)$  does not satisfy the abrupt stop threshold value, then in step 515 the speech activity is determined not to have an abrupt stop or mute.

If delta frame envelope  $\Delta e(l_M)$  does satisfy the abrupt stop threshold value, then in step 520 interval  $T_i$  is checked to determine if the speech activity is of sufficient duration, e.g., longer than a short burst. In one embodiment, the duration of interval  $T_i$  is checked to see if it exceeds the duration threshold value, e.g., 60 ms. That is, if  $T_i < 60$  ms, then the speech activity associated with interval  $T_i$  is not of sufficient duration. If the speech activity is considered not of sufficient duration, then in step 525 the speech activity is determined not to have an abrupt stop or mute.

If the speech activity is considered of sufficient duration, then in step 530 a maximum frame envelope  $e(l)$  is determined for one or more frames prior to frame  $l_M$  through frame  $l_M$  or beyond and subsequently compared against a stop-energy threshold value. The stop-energy threshold value representing a criteria for determining whether a frame envelope has sufficient energy prior to muting. In one embodiment, maximum

frame envelope  $e(l)$  is determined for frames  $l_{M-7}$  through  $l_M$  and compared to a stop-energy threshold value of 9.5, i.e.,  $\max_{l=M-7 \text{ to } l_M} e(l) > 9.5$ . If the maximum frame envelope  $e(l)$  does not satisfy the stop-energy threshold value, then in step 535 the speech activity is determined not to have an abrupt stop or mute.

5        If the maximum frame envelope  $e(l)$  does satisfy the stop-energy threshold value, then objective speech frame quality assessment  $v_s(m)$  is modified in accordance with equation 13 for several frames  $m$ , such as  $m_M, \dots, m_M+6$ :

$$\psi_s(m) = |\Delta e(l_M)| \left[ \frac{6}{1 + \exp[-2(m - m_M - 3)]} - 6 \right] \quad \text{equation (13)}$$

where  $m_M$  corresponds to the frame  $m$  which is impacted most by abrupt stop frame  $l_M$ .

10      Fig. 6 depicts a flowchart 600 illustrating an embodiment for determining whether speech activity has an abrupt start and for modifying objective speech frame quality assessment  $v_s(m)$  when it is determined that such speech activity has an abrupt start. In step 605, abrupt start frame  $l_S$  is determined. The abrupt start frame  $l_S$  is determined by first finding positive peaks of delta frame envelope  $\Delta e(l)$  in the speech activity using all frames  $l$  in interval  $T_i$ . Delta frame envelope  $\Delta e(l)$  has a positive peak at  $l$  if  $\Delta e(l) > \Delta e(l+j)$  for  $3 \leq j \leq 3$ . Upon finding the positive peaks, abrupt start frame  $l_S$  is determined as the maximum of the positive peaks of delta frame envelopes  $\Delta e(l)$ . In step 15      610, delta frame envelope  $\Delta e(l_S)$  is checked to determine whether an abrupt start threshold value is satisfied. The abrupt start threshold representing a criteria for determining whether there was sufficient positive change in frame envelope from one frame  $l$  to another frame  $l+1$  to be considered an abrupt start. In one embodiment, the abrupt stop threshold value is 0.9 and step 610 may be expressed as equation (14):

$$\Delta e(l_S) > 0.9 \quad \text{equation (14)}$$

If delta frame envelope  $\Delta e(l_S)$  does not satisfy the abrupt start threshold value, then in 20      25 step 615 the speech activity is determined not to have an abrupt start:

If delta frame envelope  $\Delta e(l_S)$  does satisfy the abrupt start threshold value, then in step 620 interval  $T_i$  is checked to determine if the speech activity is of sufficient duration, e.g., longer than a short burst. In one embodiment, the duration of interval  $T_i$  is checked to see if it exceeds the short burst threshold value, e.g., 60 ms. That is, if  $T_i < 60$

ms, then the speech activity associated with interval  $T_i$  is not of sufficient duration. If the speech activity is not of sufficient duration, then in step 625 the speech activity is determined not to have an abrupt start.

If the speech activity is of sufficient duration, then in step 630 a maximum frame envelope  $e(l)$  is determined for frame  $l_S$  or prior through one or more frames after frame  $l_S$  and subsequently compared against a start-energy threshold value. The start-energy threshold value representing a criteria for determining whether a frame envelope has sufficient energy. In one embodiment, maximum frame envelope  $e(l)$  is determined for frames  $l_S$  through  $l_S + 7$  and compared to a start-energy threshold value of 12, i.e.,

5      max <sub>$l_S \leq l \leq l_S + 7$</sub>   $e(l) < 12$ . If the maximum frame envelope  $e(l)$  does not satisfy the start-energy threshold value, then in step 635 the speech activity is determined not to have an abrupt start.

If the maximum frame envelope  $e(l)$  does satisfy the start-energy threshold value, then objective speech frame quality assessment  $v_s(m)$  is modified in accordance with equation 16 for several frames  $m$ , such as  $m_M, \dots, m_M + 6$ :

$$\%_g(m) = \frac{v_s(m)}{1 + \exp[-0.4(m - m_S)/\Delta e(l_S) - 10]} \quad \text{equation (16)}$$

where  $m_S$  corresponds to the frame  $m$  which is impacted most by abrupt start frame  $l_S$ . It should be understood that the values used in equations (11), (13) and (16) were derived empirically. Other values are possible. Thus, the present invention should not be limited to those specific values.

Note that upon determining modified objective speech frame quality assessment  $\%_g(m)$ , the integration performed in step 145 may be achieved using equation (17):

$$v_s(m) = \min(v_{s,I}(m), v_{s,M}(m), v_{s,S}(m)) \quad \text{equation (17)}$$

25 where  $v_{s,I}(m)$ ,  $v_{s,M}(m)$  and  $v_{s,S}(m)$  correspond to the modified objective speech frame quality assessment  $\%_g(m)$  of equations 11, 13 and 16, respectively.

Although the present invention has been described in considerable detail with reference to certain embodiments, other versions are possible. For example, the orders of the steps in the flowcharts may be re-arranged, or some steps (or criteria) may

be deleted from or added to the flowcharts. Therefore, the spirit and scope of the present invention should not be limited to the description of the embodiments contained herein. It should also be understood to those skilled in the art that the present invention may be implemented either as hardware or software incorporated into some type of processor.

Claims

I claim:

1. A method for objectively assessing speech quality comprising the steps of:
  - detecting distortions in an interval of speech activity using envelope information; and
    - modifying an objective speech quality assessment value associated with the speech activity to reflect the impact of the distortions on subjective speech quality assessment.
- 10 2. The method of claim 1, wherein the step of modifying includes the step of determining the objective speech quality assessment values for the speech activity.
- 15 3. The method of claim 1, wherein the distortions being detected are impulsive noise, abrupt stop or abrupt start.
4. The method of claim 1, wherein the step of detecting includes the step of determining a distortion type.
- 20 5. The method of claim 4, wherein the distortion type is determined to be impulsive noise if the envelope information indicates that the speech activity can be perceived by a human listener to be noise and if the interval is of a duration long enough to be perceived by a human listener but not too long for a short burst.
- 25 6. The method of claim 4, wherein the distortion type is determined to be impulsive noise if the envelope information indicates that the speech activity can be perceived by a human listener to be noise, if a ratio of the objective speech quality assessment value to a modulation noise reference unit indicates a human listener would perceive annoying noise, and if the interval is of a duration long enough to be perceived by a human listener but not too long for a short burst.
- 30

7. The method of claim 4, wherein the objective quality assessment value associated with the speech activity is modified in accordance with the following equation to obtain a modified objective quality assessment value if the distortion type is impulsive noise:

5

$$\text{V}_p(m) = \frac{v_s(m)}{1 + \exp[-8.2(m - m_1)/e(l_1) - 10]}$$

where  $v_s(m)$  is the objective quality assessment value and  $\text{V}_p(m)$  is the modified objective quality assessment value.

8. The method of claim 4, wherein the distortion type is determined to be abrupt stop  
10 if the envelope information indicates that there was an sufficient negative change in frame energy from one frame to another to be considered an abrupt stop and if the interval is of a duration longer than a short burst.

9. The method of claim 4, wherein the distortion type is determined to be abrupt stop  
15 if the envelope information indicates that a maximum frame envelope had sufficient energy prior to ending the interval, and if the interval is of a duration longer than a short burst.

10. The method of claim 4, wherein the objective quality assessment value associated  
20 with the speech activity is modified in accordance with the following equation to obtain a modified objective quality assessment value if the distortion type is impulsive noise:

$$\text{V}_p(m) = |\Delta e(l_M)| \left[ \frac{6}{1 + \exp[-2(m - m_M - 3)]} - 6 \right]$$

25 where  $v_s(m)$  is the objective quality assessment value and  $\text{V}_p(m)$  is the modified objective quality assessment value.

11. The method of claim 4, wherein the distortion type is determined to be abrupt start if the envelope information indicates that there was an sufficient positive change

in frame energy from one frame to another to be considered an abrupt start and if the interval is of a duration longer than a short burst.

12. The method of claim 4, wherein the distortion type is determined to be abrupt stop if the envelope information indicates that a maximum frame envelope had sufficient energy towards a beginning of the interval, and if the interval is of a duration longer than a short burst.
13. The method of claim 4, wherein the objective quality assessment value associated with the speech activity is modified in accordance with the following equation to obtain a modified objective quality assessment value if the distortion type is impulsive noise:

$$\%_o(m) = \frac{v_s(m)}{1 + \exp[-0.4(m - m_s)/\Delta e(l_s) - 10]}$$

where  $v_s(m)$  is the objective quality assessment value and  $\%_o(m)$  is the modified objective quality assessment value.

14. The method of claim 1 comprising the additional step of:  
prior to the step of detecting, determining the interval of speech activity using the envelope information.
15. An objective speech quality assessment system comprising:  
means for detecting distortions in an interval of speech activity using envelope information; and  
means for modifying an objective speech quality assessment value associated with the speech activity to reflect the impact of the distortions on subjective speech quality assessment.
16. The objective speech quality assessment system of claim 15, wherein the means for modifying includes a means for determining the objective speech quality assessment values without accounting for distortions for the speech activity.

17. The objective speech quality assessment system of claim 15, wherein the distortions being detected are impulsive noise, abrupt stop or abrupt start.
- 5 18. The objective speech quality assessment system of claim 15, wherein the means for detecting includes a means for determining a distortion type.
- 10 19. The objective speech quality assessment system of claim 18, wherein the means for detecting includes a voice activity detector for detecting intervals of speech activity, wherein the means for determining a distortion type examines intervals of speech activities detected by the voice activity detector.

Abstract of the Disclosure

Disclosed is an objective speech quality assessment technique that reflects the impact of distortions which can dominate overall speech quality assessment by modeling the impact of such distortions on subjective speech quality assessment, thereby, 5 accounting for language effects in objective speech quality assessment.

**COMPENSATION FOR UTTERANCE DEPENDENT ARTICULATION  
FOR SPEECH QUALITY ASSESSMENT**

**Field of the Invention**

5       The present invention relates generally to communications systems and, in particular, to speech quality assessment.

**Background of the Related Art**

Performance of a wireless communication system can be measured,  
10 among other things, in terms of speech quality. In the current art, there are two techniques of speech quality assessment. The first technique is a subjective technique (hereinafter referred to as "subjective speech quality assessment"). In subjective speech quality assessment, human listeners are used to rate the speech quality of processed speech, wherein processed speech is a transmitted speech signal which has been  
15 processed at the receiver. This technique is subjective because it is based on the perception of the individual human, and human assessment of speech quality typically takes into account phonetic contents, speaking styles or individual speaker differences. Subjective speech quality assessment can be expensive and time consuming.

The second technique is an objective technique (hereinafter referred to as  
20 "objective speech quality assessment"). Objective speech quality assessment is not based on the perception of the individual human. Most objective speech quality assessment techniques are based on known source speech or reconstructed source speech estimated from processed speech. However, these objective techniques do not account for phonetic contents, speaking styles or individual speaker differences.

25       Accordingly, there exists a need for assessing speech quality objectively which takes into account phonetic contents, speaking styles or individual speaker differences.

**Summary of the Invention**

30       The present invention is a method for objective speech quality assessment that accounts for phonetic contents, speaking styles or individual speaker differences by

distorting speech signals under speech quality assessment. By using a distorted version of a speech signal, it is possible to compensate for different phonetic contents, different individual speakers and different speaking styles when assessing speech quality. The amount of degradation in the objective speech quality assessment by distorting the 5 speech signal is maintained similarly for different speech signals, especially when the amount of distortion of the distorted version of speech signal is severe. Objective speech quality assessment for the distorted speech signal and the original undistorted speech signal are compared to obtain a speech quality assessment compensated for utterance dependent articulation. In one embodiment, the comparison corresponds to a difference 10 between the objective speech quality assessments for the distorted and undistorted speech signals.

**Brief Description of the Drawings**

The features, aspects, and advantages of the present invention will become 15 better understood with regard to the following description, appended claims, and accompanying drawings where:

Fig. 1 depicts an objective speech quality assessment arrangement which compensates for utterance dependent articulation in accordance with the present invention;

20 Fig. 2 depicts an embodiment of an objective speech quality assessment module employing an auditory-articulatory analysis module in accordance with the present invention.;

Fig. 3 depicts a flowchart for processing, in an articulatory analysis module, the plurality of envelopes  $a_i(t)$  in accordance with one embodiment of the invention; and

25 Fig. 4 depicts an example illustrating a modulation spectrum  $A_i(m,f)$  in terms of power versus frequency.

**Detailed Description**

The present invention is a method for objective speech quality assessment 30 that accounts for phonetic contents, speaking styles or individual speaker differences by distorting processed speech. Objective speech quality assessment tend to yield different

values for different speech signals which have same subjective speech quality scores. The reason these values differ is because of different distributions of spectral contents in the modulation spectral domain. By using a distorted version of a processed speech signal, it is possible to compensate for different phonetic contents, different individual speakers and different speaking styles. The amount of degradation in the objective speech quality assessment by distorting the speech signal is maintained similarly for different speech signals, especially when the distortion is severe. Objective speech quality assessment for the distorted speech signal and the original undistorted speech signal are compared to obtain a speech quality assessment compensated for utterance dependent articulation.

Fig. 1 depicts an objective speech quality assessment arrangement 10 which compensates for utterance dependent articulation in accordance with the present invention. Objective speech quality assessment arrangement 10 comprises a plurality of objective speech quality assessment modules 12, 14, a distortion module 16 and a compensation utterance-specific bias module 18. Speech signal  $s(t)$  is provided as inputs to distortion module 16 and objective speech quality assessment module 12. In distortion module 16, speech signal  $s(t)$  is distorted to produce a modulated noise reference unit (MNRU) speech signal  $s'(t)$ . In other words, distortion module 16 produces a noisy version of input signal  $s(t)$ . MNRU speech signal  $s'(t)$  is then provided as input to objective speech quality assessment module 14.

In objective speech quality assessment modules 12, 14, speech signal  $s(t)$  and MNRU speech signal  $s'(t)$  are processed to obtain objective speech quality assessments  $SQ(s(t))$  and  $SQ(s'(t))$ . Objective speech quality assessment modules 12, 14 are essentially identical in terms of the type of processing performed to any input speech signals. That is, if both objective speech quality assessment modules 12, 14 receive the same input speech signal, the output signals of both modules 12, 14 would be approximately identical. Note that, in other embodiments, objective speech quality assessment modules 12, 14 may process speech signals  $s(t)$  and  $s'(t)$  in a manner different from each other. Objective speech quality assessment modules are well-known in the art. An example of such a module will be described later herein.

Objective speech quality assessments  $SQ(s(t))$  and  $SQ(s'(t))$  are then compared to obtain speech quality assessment  $SQ_{compensated}$ , which compensates for

utterance dependent articulation. In one embodiment, speech quality assessment  $SQ_{compensated}$  is determined using the difference between objective speech quality assessments  $SQ(s(t))$  and  $SQ(s'(t))$ . For example,  $SQ_{compensated}$  is equal to  $SQ(s(t))$  minus  $SQ(s'(t))$ , or vice-versa. In another embodiment, speech quality assessment  $SQ_{compensated}$  is determined based on a ratio between objective speech quality assessments  $SQ(s(t))$  and  $SQ(s'(t))$ . For example,

$$SQ_{compensated} = \frac{SQ(s(t))+\mu}{SQ(s'(t))+\mu} \quad \text{or} \quad SQ_{compensated} = \frac{SQ(s'(t))+\mu}{SQ(s(t))+\mu}$$

where  $\mu$  is a small constant value.

As mentioned earlier, objective speech quality assessment modules 12, 14  
 10 are well known in the art. Fig. 2 depicts an embodiment 20 of an objective speech  
 quality assessment module 12, 14 employing an auditory-articulatory analysis module in  
 accordance with the present invention. As shown in Fig. 2, objective quality assessment  
 module 20 comprises of cochlear filterbank 22, envelope analysis module 24 and  
 articulatory analysis module 26. In objective quality assessment module 20, speech  
 15 signal  $s(t)$  is provided as input to cochlear filterbank 22. Cochlear filterbank 22  
 comprises a plurality of cochlear filters  $h_i(t)$  for processing speech signal  $s(t)$  in  
 accordance with a first stage of a peripheral auditory system, where  $i=1,2,\dots,N_c$  represents  
 a particular cochlear filter channel and  $N_c$  denotes the total number of cochlear filter  
 channels. Specifically, cochlear filterbank 22 filters speech signal  $s(t)$  to produce a  
 20 plurality of critical band signals  $s_i(t)$ , wherein critical band signal  $s_i(t)$  is equal to  
 $s(t)*h_i(t)$ .

The plurality of critical band signals  $s_i(t)$  is provided as input to envelope  
 analysis module 24. In envelope analysis module 24, the plurality of critical band signals  
 $s_i(t)$  is processed to obtain a plurality of envelopes  $a_i(t)$ , wherein  $a_i(t)=\sqrt{s_i^2(t)+\hat{s}_i^2(t)}$  and  
 25  $\hat{s}_i(t)$  is the Hilbert transform of  $s_i(t)$ .

The plurality of envelopes  $a_i(t)$  is then provided as input to articulatory  
 analysis module 26. In articulatory analysis module 26, the plurality of envelopes  $a_i(t)$  is  
 processed to obtain a speech quality assessment for speech signal  $s(t)$ . Specifically,  
 articulatory analysis module 26 does a comparison of the power associated with signals  
 30 generated from the human articulatory system (hereinafter referred to as "articulation

power  $P_A(m,i)$ ") with the power associated with signals not generated from the human articulatory system (hereinafter referred to as "non-articulation power  $P_{NA}(m,i)$ "). Such comparison is then used to make a speech quality assessment.

Fig. 3 depicts a flowchart 300 for processing, in articulatory analysis module 26, the plurality of envelopes  $a_i(t)$  in accordance with one embodiment of the invention. In step 310, Fourier transform is performed on frame  $m$  of each of the plurality of envelopes  $a_i(t)$  to produce modulation spectrums  $A_i(m,f)$ , where  $f$  is frequency.

Fig. 4 depicts an example 40 illustrating modulation spectrum  $A_i(m,f)$  in terms of power versus frequency. In example 40, articulation power  $P_A(m,i)$  is the power associated with frequencies 2~12.5 Hz, and non-articulation power  $P_{NA}(m,i)$  is the power associated with frequencies greater than 12.5 Hz. Power  $P_{No}(m,i)$  associated with frequencies less than 2 Hz is the DC-component of frame  $m$  of critical band signal  $a_i(t)$ . In this example, articulation power  $P_A(m,i)$  is chosen as the power associated with frequencies 2~12.5 Hz based on the fact that the speed of human articulation is 2~12.5 Hz, and the frequency ranges associated with articulation power  $P_A(m,i)$  and non-articulation power  $P_{NA}(m,i)$  (hereinafter referred to respectively as "articulation frequency range" and "non-articulation frequency range") are adjacent, non-overlapping frequency ranges. It should be understood that, for purposes of this application, the term "articulation power  $P_A(m,i)$ " should not be limited to the frequency range of human articulation or the aforementioned frequency range 2~12.5 Hz. Likewise, the term "non-articulation power  $P_{NA}(m,i)$ " should not be limited to frequency ranges greater than the frequency range associated with articulation power  $P_A(m,i)$ . The non-articulation frequency range may or may not overlap with or be adjacent to the articulation frequency range. The non-articulation frequency range may also include frequencies less than the lowest frequency in the articulation frequency range, such as those associated with the DC-component of frame  $m$  of critical band signal  $a_i(t)$ .

In step 320, for each modulation spectrum  $A_i(m,f)$ , articulatory analysis module 26 performs a comparison between articulation power  $P_A(m,i)$  and non-articulation power  $P_{NA}(m,i)$ . In this embodiment of articulatory analysis module 26, the comparison between articulation power  $P_A(m,i)$  and non-articulation power  $P_{NA}(m,i)$  is an

articulation-to-non-articulation ratio  $\text{ANR}(m,i)$ . The ANR is defined by the following equation

$$\text{ANR}(m,i) = \frac{P_A(m,i) + \epsilon}{P_{NA}(m,i) + \epsilon} \quad \text{equation (1)}$$

where  $\epsilon$  is some small constant value. Other comparisons between articulation power  $P_A(m,i)$  and non-articulation power  $P_{NA}(m,i)$  are possible. For example, the comparison may be the reciprocal of equation (1), or the comparison may be a difference between articulation power  $P_A(m,i)$  and non-articulation power  $P_{NA}(m,i)$ . For ease of discussion, the embodiment of articulatory analysis module 26 depicted by flowchart 300 will be discussed with respect to the comparison using  $\text{ANR}(m,i)$  of equation (1). This should not, however, be construed to limit the present invention in any manner.

In step 330,  $\text{ANR}(m,i)$  is used to determine local speech quality  $\text{LSQ}(m)$  for frame  $m$ . Local speech quality  $\text{LSQ}(m)$  is determined using an aggregate of the articulation-to-non-articulation ratio  $\text{ANR}(m,i)$  across all channels  $i$  and a weighing factor  $R(m,i)$  based on the DC-component power  $P_{No}(m,i)$ . Specifically, local speech quality  $\text{LSQ}(m)$  is determined using the following equation

$$\text{LSQ}(m) = \log \left[ \sum_{i=1}^{N_r} \text{ANR}(m,i) R(m,i) \right] \quad \text{equation (2)}$$

where

$$R(m,i) = \frac{\log(1+P_{No}(m,i))}{\sum_{k=1}^{N_r} \log(1+P_{No}(m,k))} \quad \text{equation (3)}$$

and  $k$  is a frequency index.

In step 340, overall speech quality  $SQ$  for speech signal  $s(t)$  is determined using local speech quality  $\text{LSQ}(m)$  and a log power  $P_s(m)$  for frame  $m$ . Specifically, speech quality  $SQ$  is determined using the following equation

$$SQ = L \left\{ P_s(m) \text{LSQ}(m) \right\}_{m=1}^T = \left[ \sum_{\substack{m=1 \\ P_s > P_a}}^T P_s^A(m) \text{LSQ}^A(m) \right]^K \quad \text{equation (4)}$$

where  $P_s(m) = \log \left[ \sum_{t=m}^L s^2(t) \right]$ ,  $L$  is  $L_p$ -norm,  $T$  is the total number of frames in speech signal  $s(t)$ ,  $\lambda$  is any value, and  $P_{th}$  is a threshold for distinguishing between audible signals and silence. In one embodiment,  $\lambda$  is preferably an odd integer value.

The output of articulatory analysis module 26 is an assessment of speech quality SQ over all frames  $m$ . That is, speech quality SQ is a speech quality assessment for speech signal  $s(t)$ .

Although the present invention has been described in considerable detail with reference to certain embodiments, other versions are possible. Therefore, the spirit and scope of the present invention should not be limited to the description of the 10 embodiments contained herein.